

# REPORT DOCUMENTATION PAGE

AFRL-SR-BL-TR-98-

Public reporting burden for this collection of information is estimated to average 1 hour per response, including gathering and maintaining the data needed, and completing and reviewing the collection of information. Send collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Project, Washington, DC 20503.

Source:  
of this  
Person

0289

1. AGENCY USE ONLY (Leave blank)

2. REPORT DATE

March 15, 1998

3. REPORT TYPE AND DATES COVERED

Final Technical 05/01/97 - 01/31/98

4. TITLE AND SUBTITLE

Advanced Automatic Target Recognition

5. FUNDING NUMBERS

F 49620-97-1-0523

6. AUTHOR(S)

Rama Chellappa

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Center for Automation Research  
University of Maryland  
College Park, MD 20742

8. PERFORMING ORGANIZATION  
REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Air Force Office of Scientific Research  
110 Duncan Avenue, Suite B115  
Bolling Air Force Base 20332

10. SPONSORING/MONITORING  
AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

19980414 135

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release.  
Distribution unlimited.

12b. DISTRIBUTION STATEMENT

13. ABSTRACT (Maximum 200 words)

This final report summarizes the findings of the research, "Advanced Automatic Target Recognition", supported by AFOSR grant F49620-97-1-0523. In this research effort, we have developed a method for detecting buildings from SAR images, so that false alarms due to building returns can be reduced. We consider three scenarios corresponding to incremental data availability from a high-resolution, airborne SAR, multiple SAR images and interferometric SAR. In a single strip-map SAR image, we look for certain characteristics exhibited by buildings in radar imagery, namely the combination of cardinal streaks and supporting shadow, to delineate buildings. We then present a framework for registering multi-pass airborne SAR images and for extracting heights of 3-D structures which produce identifiable linear patterns in them. Finally, given noisy elevation data derived from an interferometric (IF) SAR, buildings are segmented from the ground using a local histogram-based thresholding scheme, consolidated by propagating the thresholds, and refining along their edges to reduce errors. The effectiveness of the building detection and height estimation algorithms is demonstrated using examples of high-resolution SAR data from Lincoln Laboratory's ADTS radar and elevation data derived from Sandia's IFSAR platform. Our results will make possible on-the-fly, context-based exploitation of SAR images.

14. SUBJECT TERMS

Building detection, height estimation, interferometric SAR.  
SAR images

15. NUMBER OF PAGES

16. PRICE CODE

17. SECURITY CLASSIFICATION  
OF REPORT

UNCLASSIFIED

18. SECURITY CLASSIFICATION  
OF THIS PAGE

UNCLASSIFIED

19. SECURITY CLASSIFICATION  
OF ABSTRACT

UNCLASSIFIED

20. LIMITATION OF ABSTRACT

UL

# 1 Summary

This final report summarizes the findings of the research, "Advanced Automatic Target Recognition", supported by AFOSR grant F49620-97-I-0523. One of the difficult problems in detecting and recognizing targets in synthetic aperture radar (SAR) imagery taken over built up areas is the enormous number of false alarms due to bright returns from buildings and other man-made scatterers. In this research effort, we have developed a method for detecting buildings from SAR images, so that false alarms due to building returns can be reduced. The method combines statistical distribution of intensity and geometric attributes of man-made objects in extracting buildings.

We consider three scenarios corresponding to incremental data availability from a high-resolution, airborne SAR, and develop building-detection and/or height extraction algorithms for each. In a single strip-map SAR image, we look for certain characteristics exhibited by buildings in radar imagery, namely the combination of cardinal streaks and supporting shadow, to delineate buildings. Although building heights can be estimated from shadow extents in monocular radar images, matching rooftop features across multiple images provide better estimates. We then present a framework for registering multi-pass, airborne SAR images and for extracting heights of 3-D structures which produce identifiable linear patterns in them. Finally, given noisy elevation data derived from an interferometric (IF) SAR, buildings are segmented from the ground using a local histogram-based thresholding scheme, consolidated by propagating the thresholds, and refining along their edges to reduce errors. The effectiveness of building detection and height estimation algorithms is demonstrated using examples of high-resolution SAR data from Lincoln Laboratory's ADTS radar and elevation data derived from Sandia's IFSAR platform.

Our results make possible on-the-fly, context-based exploitation of SAR images.

## **Publications Resulting from this Contract:**

1. S. Kutikkad and R. Chellappa, "Detecting buildings in imagery derived from airborne SAR data" submitted for Computer Vision and Image Understanding, Academic Press.

## **Graduate Student Supported:**

Mr. Shaym Kuttikkad supported by this contract successfully defended his Ph.D. thesis in December 1997.

# 2 Introduction

Buildings are the dominant structures in most high-resolution aerial imagery of urban areas. Building-detection finds application in several military tasks, such as surveillance, monitoring, targeting, mission planning, training, and damage assessment. Civilian uses for this capability can be found in cartography, construction of geospatial databases, land-use surveying, and urban planning. Despite the ease with which humans can recognize buildings in aerial images, automatic

detection and height estimation of buildings is a difficult task. Illumination variations, occlusions, variations in the material and color of roofs and walls, and the assortment of shapes and sizes of buildings, add to the complexity of the problem.

Work on detecting buildings in monocular electro-optical (EO) images has mostly focussed on detecting edges and corners, grouping them by fitting polygonal sides, and subsequent high-level reasoning, possibly incorporating shadow evidence [13, 14, 27, 32, 25, 23, 26, 24]. The projected dimensions of the vertical sides of buildings have been used to aid detection and to compute heights in oblique views [12, 26, 24]. Stereoscopic building-height extraction by matching features across multiple images has also been demonstrated for optical images [12, 27, 31, 28, 15]. Some previous work has been done in the area of detecting buildings purely in high resolution digital elevation data. This includes elevation data derived from stereoscopic optical imagery [6, 2], laser scanners [33], and IFSAR [15, 8]. In this contract, we focus our attention on building extraction from high-resolution, airborne SAR imagery.

Unlike optical or infrared sensors, SAR is an active sensor, making it suitable for day-night operation independent of illumination conditions [3]. When designed to operate at the appropriate frequency, it is capable of penetrating cloud-cover, fog and smoke, which makes it very useful as an all-weather aerial imaging system. In this contract, we concentrate on analyzing images acquired from an airborne SAR operating in the strip-map mode. In this mode, the sensor platform maintains a constant heading, with the antenna pointed sideways and downward, sweeping its footprint along the flight path. The image thus formed is a projection of the 3-D world onto the slant-range plane. This is equivalent to an orthographic image produced by an optical sensor located along an imaginary line orthogonal to the SAR line-of-sight, illuminated by a light source at the SAR platform location. Thus, in order to obtain a near-nadir image of a site, the sensor must operate at low depression angles. Unlike in optical images, the SAR image pixel resolution on the ground is independent of distance from the sensor. In other words, the image of an object in far-range appears as large as that of a similar-sized object in near-range. These features make SAR suitable for operation from a long standoff distance, and hence a superior airborne imaging system.

A characteristic feature of SAR images is the phenomenon of speckle, which manifests itself as multiplicative noise [11]. It arises from the constructive and destructive interferences which occur at the receiver of a coherent sensor, and gives rise to the grainy appearance of radar imagery. As a consequence, several low-level feature detection techniques used for optical image analysis, such as gradient-based edge detection and gray-level-based segmentation, cannot be directly applied to SAR imagery. Slant-plane imaging produces another couple of effects peculiar to SAR images: foreshortening and layover. Foreshortening leads to the compressed appearance of hill-sides facing the sensor whereas layover causes the top of vertical structures to appear nearer in range than their base, in effect flipping them over. Due to its coherent sensing mechanism, SAR data acquisition and processing is expensive, requiring precision stabilization of the sensor platform and/or intelligent post-processing to remove the effects of platform instability. These factors, combined with the non-literal appearance of a SAR image (from a conventional image-processing standpoint), necessitate

specialized automatic processing techniques for SAR imagery.

In this contract, we consider three scenarios of SAR data availability, and present algorithms for building detection and/or building height extraction for each. In Section 3, we present the simplest scenario, namely the availability of a single strip-map image along with some of the radar acquisition parameters, such as depression angle and resolution. We use a combination of bright-pixel detection, maximum likelihood segmentation, and post-segmentation grouping to arrive at building hypotheses. In Section 4, we present a framework for registering two airborne radar images acquired from arbitrary flight paths. We show that the disparity in the locations of roof edges facing the sensor in the registered images can then be used to compute the approximate heights of buildings. Finally, we investigate the problem of building detection in noisy elevation data derived from an interferometric SAR. The method presented in Section 5 incorporates a local-histogram-based thresholding to extract building edges, consolidation of detections, and boundary refinement. At the end of each section, we present examples of the building extraction algorithms, applied to real high-resolution SAR data. It should be noted that the algorithms described here assume that all the data processing for SAR image formation, and height extraction from IFSAR, have been done under ideal conditions. Besides, since platform dynamics during data collection and raw phase-history data were not made available to us, we do not attempt to correct, or compensate for, deviations from the ideal scenario.

### 3 Detecting Buildings in a Single Strip-map SAR Image

SAR backscatter exhibits a large dynamic range with considerable signal-strength variability within the same class of object being imaged. When combined with the presence of speckle, this limits the utility of gradient- or gray-level-based low-level processing for SAR image analysis. Several filters have been proposed for speckle reduction with the primary intent of making visual interpretation of SAR images easy [10, 22]. They are, in general, low-pass filtering operations which result in loss of spatial and/or radiometric resolution. Although statistical methods can be used to segment SAR images, while preserving resolution [17, 9], maximum likelihood or Bayesian segmentation techniques require examples of the classes or objects they are designed to detect, a priori. Due to the lack of a single exemplar which represents all buildings uniformly, statistical segmentation alone is not sufficient for building detection.

The radar backscatter energy from an object depends on its shape as well as composition. The shape of a building, the material and shape of the roof, presence of substructures on the rooftop, shadowing and occlusion, complicate the task of segmentation algorithms. Besides, since foreshortening and layover make vertical walls invisible, their presence cannot be counted upon for building identification. But if the sensor is operating at moderate depression angles (which it will, if designed for surveillance), buildings cast shadows. Shadows in radar imagery represent a lack of signal, either due to an occluding object or due to lack of backscatter. Shadows are cast by most vertical structures, such as buildings and trees, as well as specular surfaces oriented away



from the sensor, such as calm bodies of water. Another characteristic feature of buildings is the presence of cardinal streaks. These are bright pixels forming a thick linear cluster along the roof edge closest to the radar. Cardinal streaks are the result of high backscatter from a combination of several phenomena: the roof-wall edge, layover of the vertical wall closest to the sensor, and the dihedral effect produced by the building wall and the ground. But bright streaks are also produced by railroad tracks, powerline cables, vehicles, or sometimes, poorly focussed bright point-scatterers.

Our building detection scheme looks for the occurrence of a cardinal streak and a shadow region, down-range from it, within close proximity. Bright pixels are detected using Constant False Alarm Rate (CFAR) processing [7] and shadows are identified by maximum likelihood segmentation. The remainder of this section explains how we detect these features and how we discriminate buildings from other objects which exhibit similar characteristics. A rudimentary performance validation scheme using manually registered EO imagery is also presented at the end of this section.

### 3.1 Statistical models for SAR data and bright-pixel detection

In a single polarization SAR image, the measured signal at a pixel is a vector sum of backscatter from a multitude of individual scatterers, and consists of in-phase and quadrature components. Under assumptions of fully developed speckle, i.e. the surface is rough compared to the radar wavelength, the complex return at a pixel can be approximated as a circularly symmetric Gaussian random variable, due to the Central Limit Theorem [11]. The corresponding intensity or power can be shown to be exponentially distributed. Empirical measurements on vegetation showed a departure from the complex Gaussian (Rayleigh magnitude) assumptions at high resolutions and low-to-medium depression angles. The K family of distributions has been suggested as an alternate model for the resulting spiky clutter [34]. The K distribution arises when the clutter amplitude in a cell exhibits rapid Rayleigh fluctuations, whose mean varies slowly over time, according to the Gamma distribution. Fully polarimetric SAR data has been modeled as a complex Gaussian vector with a polarimetric covariance matrix [17]. Depending upon the data type (polarimetric or single-polarization, complex or intensity, etc.) as well as the empirical fit to measured data, one of the above statistical models can be used to characterize SAR data. Our building detection algorithm does not rely heavily on any particular choice of clutter statistical model. The only requirement is that patches of sample data, representing expected terrain feature classes, be identified a priori.

CFAR processing is useful for detecting strong reflectors embedded in spatially non-homogeneous background clutter. In this pixel-based method, the signal at the cell under test is compared to an adaptive threshold, generated from a moving window of reference cells from the background. The reference cells are used to obtain estimates of the parameters of the underlying clutter statistical distribution. The adaptive threshold is a function of the desired probability of false-alarm, the test-statistic from the reference window, and the size of the reference window. In an Order Statistic (OS) CFAR detector, an order-statistic of the cells in the reference window is used to compute the threshold [30]. This detector is a good trade-off between simplicity and robustness in situations where the reference window overlaps an extended or multiple targets. For the OS CFAR processor

(based on a single ranked sample), the test statistic is the  $k$ th order-statistic from  $M$  reference cells. The cumulative distribution function (CDF) of the  $k$ th order-statistic is [4]

$$F_k(u) = k \binom{M}{k} f_0(u) F_0(u)^{k-1} [1 - F_0(u)]^{M-k}$$

where  $u$  is the test-cell amplitude, and  $F_0(\cdot)$  and  $f_0(\cdot)$  are the univariate clutter-only (null hypothesis) CDF and probability density function (PDF), respectively. The CFAR threshold can be formulated as a factor,  $\tau$ , multiplying the test statistic,  $z$ . Then the probability of false-alarm,  $P_{FA}$ , and the threshold multiplier can be related by

$$P_{FA} = \int_0^\infty Pr_0[\pi_0 \geq \tau z] f_Z(z) dz$$

where  $\pi_0$  is the voltage of the cell under test,  $f_Z(z)$  is the PDF of  $Z$ , and  $Pr_0[\cdot]$  is the probability under the null hypothesis. Once again, the choice of a clutter model (K, Rayleigh, Weibull, etc.) can be governed by empirical fit to data or the complexity of the resulting CFAR detector.

### 3.2 Building detection algorithm

In [18, 19], we developed algorithms for constructing 2-D site models from single- and fully-polarimetric SAR data. In this contract, we restrict our attention to aspects of site modeling related to building detection. Our technique is suitable for detecting simple buildings with adjacent walls perpendicular to one another. To reiterate, we are looking for the combination of cardinal streaks and shadow evidence in our search for buildings. CFAR detection produces clusters of bright pixels, some of which constitute the cardinal streaks of buildings in the image. Shadow regions, along with other expected terrain feature classes, are delineated using maximum likelihood (ML) segmentation, after a supervised selection of training areas for parameter estimation. Although the segmentation results are sensitive to the selected training areas, this is not very restrictive, since for our purposes, we can assume that the characteristics of the various classes are reasonably stationary during a particular data collection run. Although maximum a posteriori estimation has been suggested for producing smoother segmentation maps [5, 29], for our purpose of extracting shadow regions, the additional computational burden is not justified.

Most of our experiments were conducted on one-foot resolution data, which dictated the choice of size- and distance-related heuristics in our algorithm. Our scheme for building detection in a single SAR image is outlined below:

- **Bright pixel detection**

Bright pixels in a radar image may correspond to vehicles, metallic objects, cardinal streaks of buildings, or false alarms. A two-stage OS CFAR detector is used to extract the full extent of bright-pixel clusters, while keeping the number of false-alarms low. During the first pass, CFAR processing is done at each pixel with a low false-alarm probability, typically  $10^{-3}$ . A second CFAR processing step, with a higher false alarm probability (typically  $10^{-2}$ ),

is performed only in the immediate neighborhood ( $5 \times 5$ ) of previously detected pixels. A hollow square reference window with outer dimension twenty-five pixels and thickness of a single pixel was used typically. This was done to ensure a sufficient number of background pixels for robust clutter-parameter estimation, and so that, when centered on a typical-sized vehicle, no part of the vehicle would overlap with the window.

- **Supervised ML segmentation**

Training areas are used to estimate the parameters describing the statistics of the likely terrain classes. We train on nearly homogeneous patches of clearing, grass, road, shadows and trees, selected elsewhere in the data sequence. Although water bodies were expected to be among the terrain features imaged, we did not include it as one of our expected ML labels. This is because the lower amounts of backscatter from the surface of calm water makes its radar signature close to that of radar shadow. The ML estimate of the region label for each pixel is obtained by maximizing the appropriate joint conditional density function, given the region label, over a local ( $3 \times 3$ ) neighborhood. The choice of a joint spatial density function is governed by the need for a smoother segmentation map. Bright pixels detected by CFAR processing are neither re-labeled nor included in the joint distribution formulation for classifying their neighboring pixels. This is to prevent bright pixels from affecting the labels of their neighbors.

- **Building detection**

Bright pixels detected by CFAR processing are grouped into clusters by performing a morphological closing operation, followed by a connected-components labeling algorithm. Only pixels initially detected by CFAR processing are retained in each connected component. A user-defined threshold (twenty pixels, for the examples in this contract) is used to eliminate small clusters. If a building is not oriented parallel to the aircraft flight path, it is possible that it produces an L-shaped cardinal feature instead of a streak. To account for this phenomenon, a robust median line is fit to each cluster separately. The clusters for which the normalized mean absolute deviation of the fitted line exceeds a nominal threshold (3.5 gave the best results in our trials), are flagged as a possible L-shaped feature. The Hough transform technique for detecting lines is used to split such a cluster into two, while imposing the constraint that the two peaks in the transform domain correspond to nearly orthogonal lines. Since the buildings are expected to have simple shapes, their walls are assumed to be aligned along two, nearly orthogonal directions. The orientations of the surviving clusters are determined by minimizing their moment of inertia. Next, we determine the smallest rectangle with a minimum aspect ratio (two, in our trials), the same orientation as the pixel cluster, and including most of the pixels in the cluster. Not requiring that all pixels belonging to a cluster be located inside this rectangle, provides the capability to reject outliers. Blob-shaped clusters (as opposed to streaks) are rejected by thresholding on the fraction of CFAR-detected pixels in the enclosing rectangle.

Once streaks have been identified, a search is performed perpendicular to the streak and

down-range from the sensor for shadow regions in the ML labeled image. Since the average shadow and road backscatter intensities are close, it is conceivable that some of the pixels in building shadow areas have been incorrectly labeled as road, especially along the shadow boundaries. We adopt dual thresholds - minimum shadow support and minimum road-plus-shadow support - to hypothesize a building. In other words, we look for a minimum number of the pixels in a streak to be bounded down-range by either a road or a shadow pixel, of which a certain fraction have to be shadow pixels. For each pixel in a detected bright streak, we search for a supporting dark pixel as far away from it as the streak length. Although this imposes an arbitrary limit on the width of the projected rooftop as a function of the building length (or width), in the absence of such a limit streaks with shadow regions far away from them may result in false detections. The width of the rooftop of a detected building is computed to be the mean distance from its cardinal streak to its supporting shadow region. Finally, the rectangle of length equal to the streak length and width equal to the mean streak-to-shadow distance is marked as a building.

### 3.3 Experimental results, validation, and discussion

Lincoln Laboratory's ADTS SAR operates at a frequency of 33GHz producing complex, polarimetric data at a resolution of  $1' \times 1'$ . Some examples of building detection on strip-map SAR imagery from this sensor are shown in Figures 1 and 2. Figure 1(a) shows the original SAR image with near-range at the top. Figure 1(b) shows the result of two-pass OS CFAR processing for bright pixel detection. The shadow (black) and road (gray) regions in the output of ML labeling, under a polarimetric complex Gaussian model, are shown in Figure 1(c). The detected buildings are shown in Figure 1(d). Another image with larger buildings in the clear and several metallic objects nearby, is shown in Figure 2(a). The labeled image with bright-pixels, shadows and roads, and the image with detected buildings are shown in Figures 2(b) and (c) respectively.

Despite the lack of knowledge of the exact sizes and locations on the ground of the buildings in our images, we have made some attempts to validate our building detection algorithm for a single SAR image. We had optical coverage photographs for some of the radar data without camera parameters or ground-control information. A set of corresponding points and linear features were manually identified in the EO and radar images. In order to ensure that layover and foreshortening do not influence parameter estimation, the selected features had to lie on or near the ground plane. The parameters for the transformation required to register the two images were estimated from them, in a least squares sense. Due to the differences in imaging geometry, a full projective transformation is needed to register the two images. The dissimilar nature of the imagery, insufficient identifiable feature-correspondences, and errors in the manual selection of correspondences resulted in excessive distortion with the projective transformation model. Therefore we chose an affine model to register the EO image to the SAR image coordinates. This is justified, since near-nadir EO images can be considered to be orthographic projections of the 3-D scene onto the ground plane, and since the SAR imaging can be approximated by an orthographic projection as mentioned in Section 2.

Once the EO image was warped to the SAR image coordinates, approximate building footprints in it were identified manually. Determining the accuracy of our building delineation in the radar image, using a measure of the area of overlap with the building footprints in the EO image, was hampered by the approximations in registration and the layover of buildings in the radar image. Therefore, we restricted ourselves to just checking for an overlap between a detected building in the radar image and a building footprint in the EO image. Figure 1(e) shows a coverage optical image warped to correspond to the SAR image of Figure 1(a). The building detection validation results are displayed in Figure 1(f) as an image, with the correct detections and missed footprints overlaid with the false alarms. The one building in left-center which is completely missed by our algorithm is due to occlusion by a tree shadow. The two other missed buildings in the far-left of the image are due to image edge-effects, which result in truncated cardinal streaks. All other buildings in the scene are detected, although some of them are only partially extracted due to incomplete streak identification. The one false detection is produced by a linear cluster of bright pixels in foliage. For the radar image in Figure 2(a), all the buildings were detected and Figure 2(d) is provided for visual comparison.

## 4 Building Height Extraction from Multi-pass SAR Imagery

In the previous section, we dealt with detecting buildings in a single SAR image and ignored the issue of extracting its height or 3-D shape. It is possible to estimate the height of a building in a single image by computing the length of its shadow, and extrapolating from the acquisition parameters. Shadow-based height estimation is not reliable because of the difficulty in isolating the leading edge of a building's shadow region, as well as the problems that arise when shadows from multiple buildings or trees overlap. In this section, we present a more robust height extraction scheme using disparity measurements of rooftop features. The features we are interested in, are the building cardinal streaks visible in both images.

We consider the problem of computing heights of buildings from a pair of non-interferometric SAR images from the same aircraft-mounted sensor, collected from arbitrary flight paths. This requires a preliminary registration of the two images, the necessary transformation for which is derived in the next sub-section. We do not claim that this transformation is sufficient to achieve a complete pixel-level registration of two SAR images. We are merely interested in approximately registering higher level features (cardinal streaks, in this case) across multiple images.

Two SAR images of a site are projections of the 3-D surface onto different, possibly non-parallel, slant planes. Hence, a Euclidean or similarity transformation is not sufficient to register the two images. With certain assumptions (to be elaborated upon later), we show that registration of two airborne SAR images can be approximated by an affine transformation. This transformation results from a cascade of the following steps: projection of the first image to the ground plane; rotation and translation within the ground plane; and projection to the slant plane of the second image. This transformation can be derived from the sensor acquisition parameters



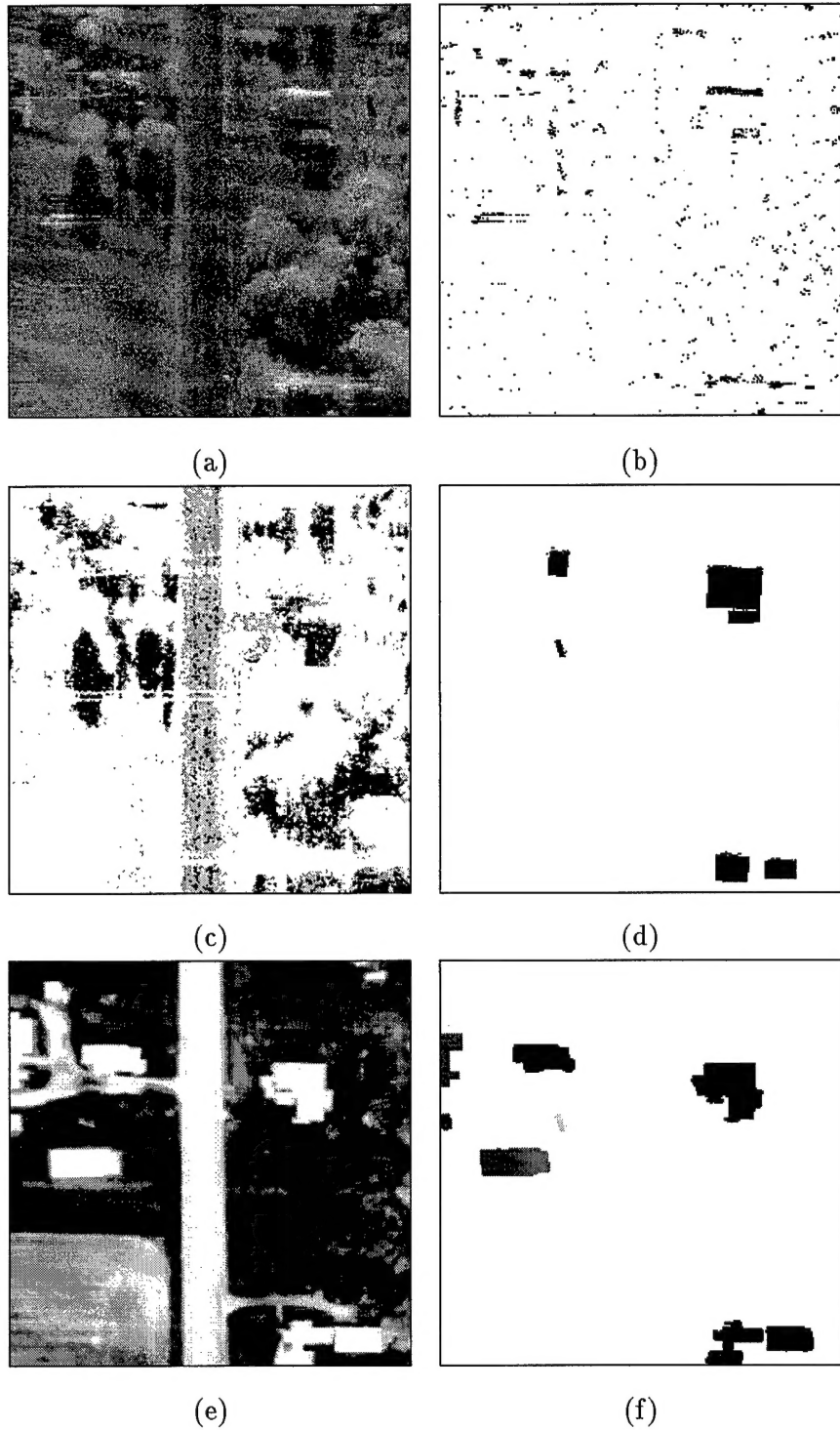
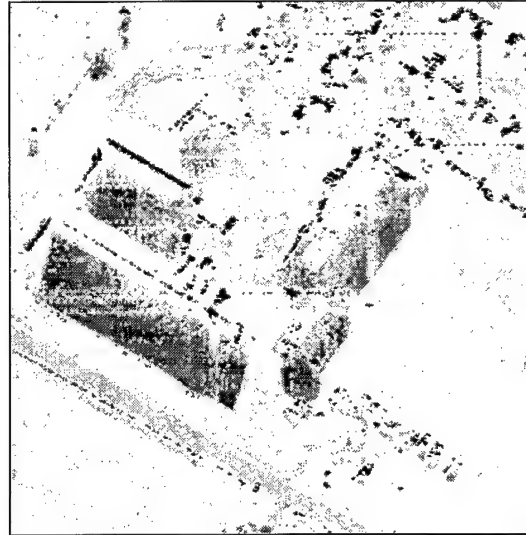


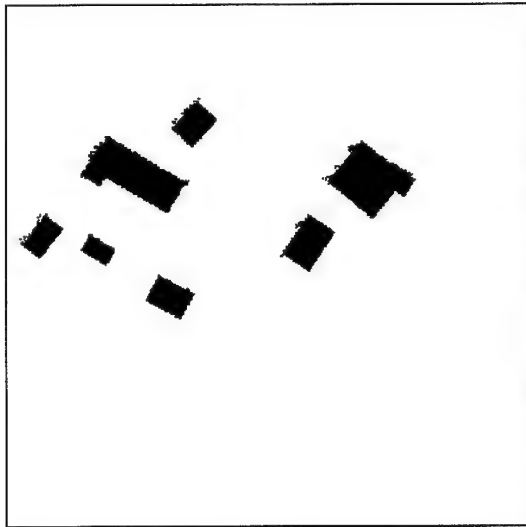
Figure 1: Building detection and validation in single-pass imagery: (a) Original image, (b) CFAR detector output, (c) Shadows (black) and roads (gray), (d) Detected buildings, (e) Registered optical image, (f) Validation results (black=detections, dark gray=misses, light gray=false alarms)



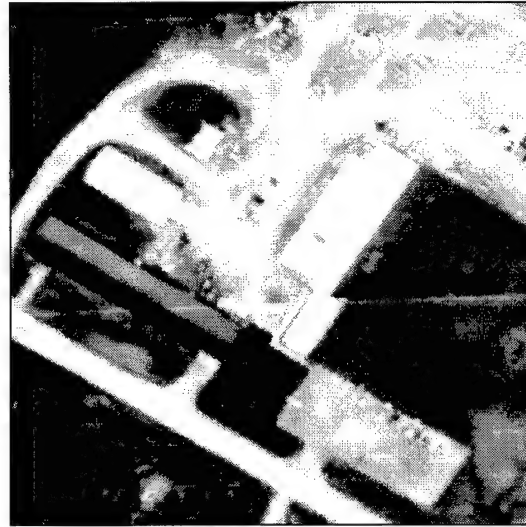
(a)



(b)



(c)



(d)

Figure 2: Another example of building detection in single-pass imagery: (a) Original image, (b) ML labeled image (black=CFAR detections, dark gray=shadow, light gray=road), (c) Detected buildings, (d) Registered optical image

#### 4.1 Registration of airborne SAR images

Figure 3 illustrates the geometry of an airborne SAR system operating in the strip-map mode. The  $XGZ$  coordinate system (represented by unit vectors,  $\hat{x}$ ,  $\hat{g}$ , and  $\hat{z}$  in Figure 3) has its origin in the ground plane, directly below the instantaneous position of the aircraft, flying at altitude  $z$ , with the  $X$ -axis aligned with the aircraft heading. The antenna is oriented at a depression angle  $\theta$ , downward from a plane parallel to the ground ( $X-G$ ) plane, at height  $z$ . Slant range,  $r$ , is along the direction of radiation propagation. The radar image is formed in the slant ( $X-R$ ) plane (represented by the unit vectors  $\hat{x}$  and  $\hat{r}$ ), where the range ( $R$ ) coordinate represents distance from the sensor and the cross-range or azimuth ( $X$ ) coordinate represents distance along the flight path. Ground range,  $g$ , is the projection of the slant range onto the ground plane, and  $\phi$  is the azimuth angle of the sensor relative to an arbitrarily fixed direction in the ground plane. Although the slant-range to ground-range transformation is non-linear, a linear approximation can be used, provided that the distance from the sensor to the image center is much larger than the image swath, and that the depression angle is not close to ninety degrees [20]. Both these conditions are met by a high-resolution aircraft-mounted radar system.

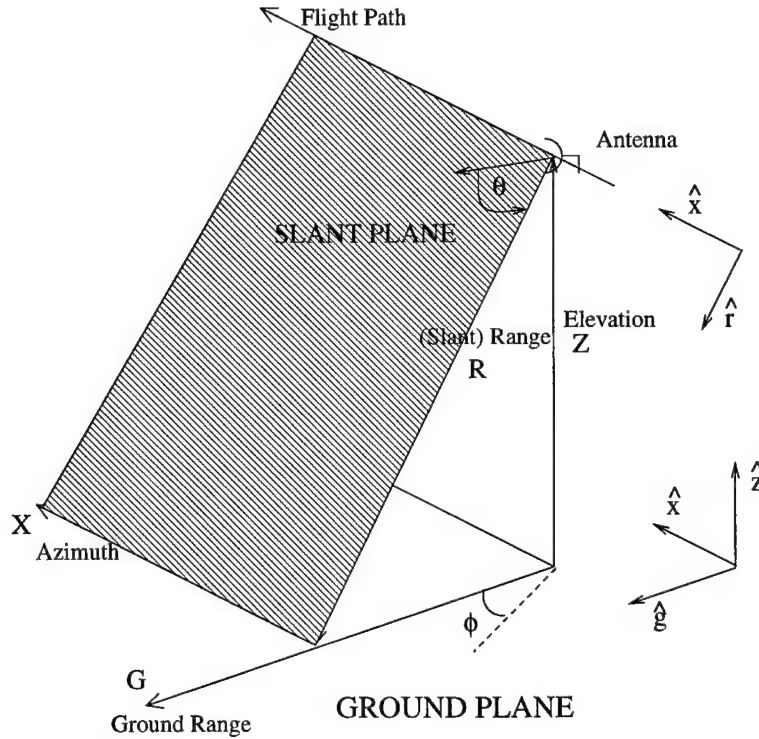


Figure 3: SAR imaging geometry

Since scatterers at a constant range from the radar are mapped into the same point, SAR image formation is a many-to-one projection of 3-D space onto a 2-D plane. While it is possible to derive the complete transformation of any 3-D point into its corresponding 2-D point in a SAR image, it is not possible to register two arbitrary SAR images without making some simplifying approximations. The first approximation is the “flat earth” assumption, i.e. the imaged terrain is assumed to be flat.

Parts of the image may have significant slope, resulting in foreshortening and possible layover (e.g., buildings, trees, mountainside), but these elements are intractable without accurate pixel-to-pixel registration. By assuming flat earth and an accurate knowledge of the acquisition geometry, it is possible to register most of the image; regions of misregistration will then correspond to structures with significant height. The second approximation is that the depression angle  $\theta$  is constant across the image swath. Once again, a large range-to-swath-width ratio justifies this approximation. This enables a single transformation to be used for all the points in the image irrespective of their range location.

A 2-D point  $\mathbf{p}^{(1)} (= [x^{(1)} \ r^{(1)}]^T)$  in the first image can be transformed into the corresponding point  $\mathbf{p}^{(2)} (= [x^{(2)} \ r^{(2)}]^T)$  in the second image via the affine transformation

$$\mathbf{p}^{(2)} = \mathbf{A}\mathbf{p}^{(1)} + \mathbf{b} \quad (1)$$

where

$$\mathbf{A} = \begin{bmatrix} \frac{1}{\delta x^{(2)}} & 0 \\ 0 & \frac{1}{\delta r^{(2)}} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & \cos \theta^{(2)} \end{bmatrix} \begin{bmatrix} \cos \phi & -\sin \phi \\ \sin \phi & \cos \phi \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & \frac{1}{\cos \theta^{(1)}} \end{bmatrix} \begin{bmatrix} \delta x^{(1)} & 0 \\ 0 & \delta r^{(1)} \end{bmatrix} \quad (2)$$

and  $\mathbf{b} = [t_x \ t_r]^T$  is the translation (azimuth and range) vector required to align the images. Here,  $\delta x^{(i)}, \delta r^{(i)}$  are the respective pixel resolutions in the azimuth and range dimensions,  $\theta^{(i)}$  are the depression angles, and  $\phi = \phi^{(2)} - \phi^{(1)}$  is the difference in sensor headings between the two images. For future reference, the affine transformation of (1) can be written in the general matrix formulation

$$\begin{bmatrix} x^{(2)} \\ r^{(2)} \\ 1 \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & t_x \\ a_{21} & a_{22} & t_r \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x^{(1)} \\ r^{(1)} \\ 1 \end{bmatrix} \quad (3)$$

All the parameters in (2) are typically available, to some degree of accuracy, for airborne SAR data. The only unknowns in (3) are  $t_x$  and  $t_r$ , in order to compute which, it is necessary to know at least one set of point-correspondences. Sometimes, Global Positioning System (GPS) information is available as a reference for the scene location. In case this information is unavailable, a point correspondence has to be found manually or by an automatic scheme. The latter method is preferred due to the difficulty in the precise manual localization of point features in SAR images.

The accuracies of the resolution and depression angle parameters are usually high, because they are part of the SAR system design. Aircraft heading information is also expected to be reasonably accurate, but the site location information is either absent or not precise enough for registration. Although GPS-derived location was available with Lincoln Lab's ADTS dataset, its accuracy was insufficient for our needs. In order to compute the translation parameter, we extract a number of point features from each image and match them across images. The features chosen should lie in (or near) the ground plane, so that there are no layover effects that would affect different views differently. They should also be easy to detect, stationary and persistent across images. We have

chosen the centroids of clusters of bright pixels as our point features. These bright returns result from metallic objects and other specular reflectors in the scene. Although the radar signatures of real-world objects depend on many factors, if visible, their centroids are good candidates for matching across images. After initial registration, translations between each feature point in one image and all feature points in the other image are computed. The overall translation is estimated to be the one that results in the maximum number of one-to-one feature-point matches.

## 4.2 Estimation of object height

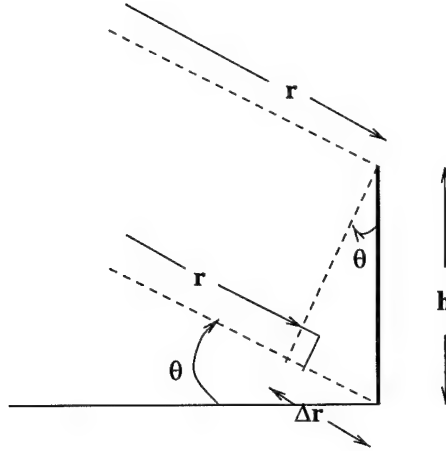


Figure 4: Observed slant range location of a tall object

Layover causes the slant plane image of the top of a 3-D structure of height  $h$  (Figure 4) to appear at a range location

$$r_{obs}^{(i)} = r^{(i)} - \Delta r^{(i)} = r^{(i)} - h \sin \theta^{(i)} \quad (4)$$

The superscript  $i$  refers to image  $i$ ,  $\theta$  is the depression angle, and  $r$  and  $r_{obs}$  are the slant range to the base and top of the structure, respectively. Mathematically, it is possible to compute the height of a tall object, given the exact location of a single point on it in two views, and the transformation between the two images. In practice, the difficulty of automatically localizing point features in SAR images and inaccurate registration lead to erroneous height computation.

Linear features, such as the cardinal streaks of buildings, are easier to detect and match. Since layover affects only the slant-range dimension, the tops of linear 3-D structures (e.g. walls) are imaged parallel to the location of their bases, but nearer in range. After undergoing an affine transformation, the top of the structure in the first image will appear parallel to, but displaced from, the same feature in the second image. We have shown in the previous section how to detect the leading roof-wall edge of buildings using CFAR processing, followed by streak extraction. Provided that the difference in azimuth between the two images is less than 90 degrees, it is conceivable that the edge of a building which produced a cardinal streak in one image is visible in the other too. After initial registration, finding correspondences between transformed and observed cardinal streaks is relatively simple, given their proximity and parallelism.



Let  $l^{(i)} (= \{p_j^{(i)}\} = \{[x_j^{(i)} \ r_j^{(i)} \ 1]^T\})$  be the location of the base of a 3-D linear structure of height  $h$  in image  $i$ . Let  $\hat{l}^{(2)} (= \{\hat{p}_j^{(2)}\})$  be the affine projection (according to (3)) of  $\tilde{l}^{(1)}$  in the second image. From (4):

$$\hat{p}_j^{(2)} = A_{(3)} \tilde{p}_j^{(1)} = A_{(3)}(p_j^{(1)} - h \sin \theta^{(1)} * [0 \ 1 \ 0]^T) \quad (5)$$

where  $A_{(3)}$  is the  $3 \times 3$  affine transformation matrix of (3). Given that the locations of points along the base of the structure in the two images are related by  $p_j^{(2)} = A_{(3)} * p_j^{(1)}$ , (5) can be rewritten as

$$\begin{aligned} \hat{p}_j^{(2)} &= p_j^{(2)} - h \sin \theta^{(1)} A_{(3)} * [0 \ 1 \ 0]^T \\ &= \tilde{p}_j^{(2)} + h(\sin \theta^{(2)} I_{(3)} - \sin \theta^{(1)} A_{(3)}) * [0 \ 1 \ 0]^T \end{aligned} \quad (6)$$

where  $I_{(3)}$  is the  $3 \times 3$  identity matrix. Substituting for  $A_{(3)}$ , (6) becomes

$$\hat{p}_j^{(2)} - \tilde{p}_j^{(2)} = \begin{bmatrix} -a_{12}h \sin \theta^{(1)} \\ h(\sin \theta^{(2)} - a_{22} \sin \theta^{(1)}) \\ 0 \end{bmatrix}$$

Although it is not possible to isolate pairs of corresponding points on the two lines, it is possible to compute height,  $h$ , by computing the perpendicular distance,  $\rho$ , between the lines in the coordinate system of the second image:

$$h = \frac{\rho}{(\sin \theta^{(2)} - a_{22} \sin \theta^{(1)}) \sin \alpha - a_{12} \sin \theta^{(1)} \cos \alpha} \quad (7)$$

where  $\alpha$  is the angle made by a line perpendicular to either of  $\hat{l}^{(2)}$  or  $\tilde{l}^{(2)}$  with the positive cross range axis in the second image.

In practice, it is difficult to ensure that the projected line  $\hat{l}^{(2)}$  is exactly parallel to the observed line,  $\tilde{l}^{(2)}$ . Since height extraction by (7) relies on the distance between two parallel line segments and their slope, refinements have to be made to the original choice of line segment locations, to make them parallel. In order to achieve this, the observed lines in the two images are de-rotated by an equal and opposite amount about their midpoints. This is justified, since the line extraction technique is the same in both images, and any errors are expected to affect both images in the same statistical sense. It is not difficult to show that the corresponding rotation angle,  $\varphi$ , is the solution of

$$\frac{a_{21} + a_{22}\tilde{m}^{(1)}}{a_{11} + a_{12}\tilde{m}^{(1)}} = \frac{-\sin \varphi + \tilde{m}^{(2)} \cos \varphi}{\cos \varphi + \tilde{m}^{(2)} \sin \varphi}$$

where  $\tilde{m}^{(i)}$  is the slope of the observed line in image  $i$ , and

$$\tilde{m}^{(1)} = \frac{\sin \varphi + \tilde{m}^{(1)} \cos \varphi}{\cos \varphi - \tilde{m}^{(1)} \sin \varphi}$$

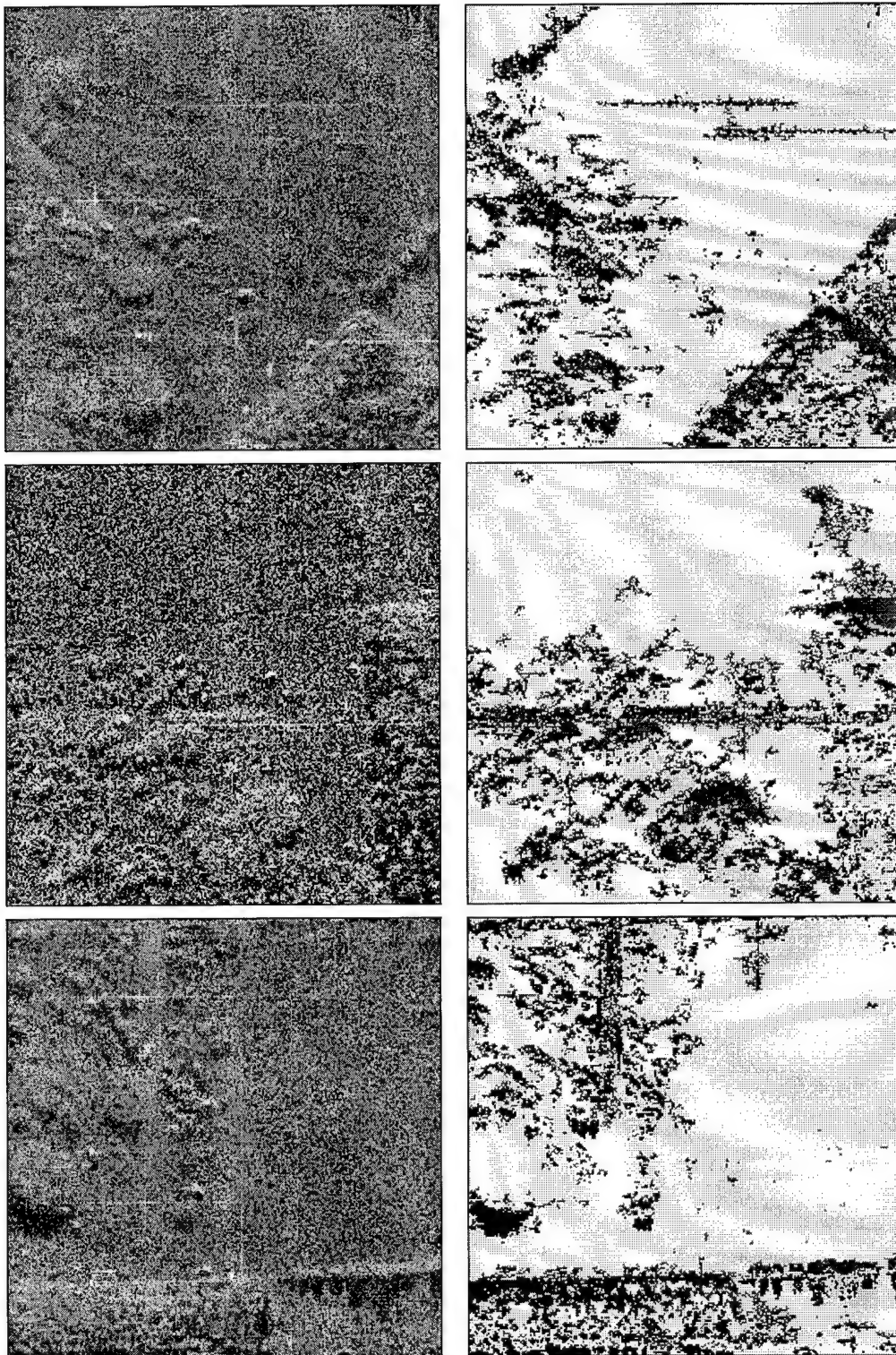


Figure 5: Tree-line height extraction: Original multi-pass images (left column) and tree-lines (red & blue lines) marked on segmented images (right column). The flight-path is across the page and range increases from top to bottom for all images.

Since we did not have real multi-pass radar imagery with buildings in them, we demonstrate the height extraction algorithm of this section by computing the height of vegetation canopy. We approximate the leading (i.e. closest to the radar) edge of short linear portions of treelines to be of constant height, and proceed to estimate this height. The process involves segmentation of each image to extract the vegetation cover, delineation of the leading edges of tree-lines, image-to-image registration, and height estimation. The left column of Figure 5 shows three views of the same site. The aircraft heading in the two lower images differed by 45 degrees to either side of the topmost image. The depression angle was approximately 23 degrees for all images, and the resolution was  $1' \times 1'$ . First, the images were segmented into four classes - trees, grass, shadows, and targets - using the bright-pixel-detection and ML-segmentation techniques described in Section 3. The resulting labeled images are shown in the right column of Figure 5 in green, yellow, black, and white, respectively. Nearly linear sections of the visible tree-lines were manually identified in the labeled images. Two such line segments (blue and red) are shown fitted to the tree/grass boundary in the top right image of Figure 5. The corresponding (visible) tree/grass edges are marked in blue in the middle image and red in the bottom image in the right column of Figure 5. After registration, the red line yielded a canopy height of 87.19 pixels ( $\approx 22\text{m}$ ), whereas the blue line yielded a height of 102.56 pixels ( $\approx 26\text{m}$ ). We are not aware of the actual height of the vegetation canopy in the imaged area, but the estimates seem reasonable for foliage in that area.

## 5 Building Detection in IFSAR-derived Elevation Data

Building-height extraction in monocular images using shadow information is unreliable because of its dependence on the imaging conditions and their accurate knowledge. Techniques based on identifying walls and extrapolating from their projected heights fail for near-nadir imagery, or under occlusion. Height computation of 3-D structures by matching features across multiple images is a difficult problem requiring accurate delineation of corresponding points and/or lines. A high-resolution grid of digital height data, on the other hand, has the building height information built into it. Of course, this requires a method for extracting terrain height reliably. Terrain height extraction from stereo reconstruction of optical binocular images requires accurate camera parameter information, as well as robust feature-extraction and matching schemes. Topography reconstruction from a stereo-pair of low resolution SAR images, acquired from the same side, opposite sides, or intersecting flight paths have been described in [21], ch. 13,14. In general, these techniques attempt to obtain a dense height map by matching regions in two images using some form of correlation-based metric.

A more robust technique for reconstructing terrain heights is to use radar interferometry from two coherently acquired images [35]. Knowing the baseline separation between the sensors during the two acquisitions, accurate phase unwrapping techniques are used to extract height information. Radar interferometry does not rely on feature correspondences and is capable of generating a dense height-map, while retaining the general advantages of a SAR imaging system over other airborne

sensors.

Earlier work on extracting buildings from high-resolution Digital Elevation Models (DEMs) derived from pairs of optical images, or laser scanners include the use of: mesh triangulation to hypothesize buildings [6], split-and-merge schemes for segmenting buildings from ground [2], and parametric and prismatic models for model-based building recognition [33]. There has been some recent work on using DEMs derived from IFSAR for building detection. In [15], the DEMs are used only to provide height samples for footprints of buildings detected in optical imagery. The authors proceed to fit one of a small set of shapes to the rooftops of detected buildings, ignoring the presence of substructures on rooftops, or multi-level buildings. In [8], an initial segmentation of the height map is refined using images from auxiliary sensors, but no details of the specific algorithm are provided.

IFSAR derived DEMs are inherently noisy and may have missing data, rendering model-based or triangulation-based building detection schemes unreliable. Besides, foreshortening and layover produce errors in the height map along building edges. In the data we experimented with, we noticed that the height estimates along the leading edges of buildings are fuzzier and less reliable than the trailing edges, due to layover. Thus, a global thresholding scheme to separate all buildings from ground will not work well either. Besides, a global scheme is bound to fail if the terrain elevation changes significantly across the image. Fitting statistical models to the data from rooftops and the ground would better describe the data. Since building heights vary and are not known a priori, building-detection via statistical model-based segmentation of elevation data is not very practical.

## 5.1 Local histogram-based thresholding for building detection

We use a local histogram-based thresholding scheme to discriminate buildings from ground in IFSAR-derived high resolution DEMs. Localized histograms have been used for segmenting gray-scale images into homogeneous regions [1]. The inherent assumption of our strategy is that the high-resolution data consists of reasonably separated buildings of possibly different heights. Although the ground elevation may change across the image, each individual building is assumed to be on a level ground. This implies that within a local neighborhood smaller than the distance between adjacent buildings, the image pixels belong to one of two classes - rooftop or ground. We assume that other tall structures in the scene (e.g., trees) can be distinguished from buildings in some other way, perhaps by segmenting the SAR data as in [18]. Due to the noisy nature of IFSAR elevation data, determining shapes of roofs is difficult and therefore we restrict ourselves to a flat-roof hypothesis.

We now present a brief overview of our method, and provide details later in this section. In order to prepare the data for thresholding we first fill in missing values and perform data smoothing. Due to its statistical nature, entropy-based measures are used to threshold non-overlapping blocks of the data. During this process, homogeneous blocks belonging either to rooftop or ground are identified. The data is re-thresholded over a larger area around certain blocks in order to correct for uncharacteristically high thresholds, resulting in rooftop pixels being labeled as part of the

ground. Thereafter, homogeneous blocks are classified as part of building or ground based on the thresholds of the blocks in their neighborhoods. Connected building regions are identified and filled in. Building edges are refined using estimates of building and ground heights from their immediate neighborhood. Finally, we look for and remove false positives with average elevation lower than that of the surrounding ground. The details of this algorithm are described next:

**Normalization and smoothing:** Elevation data derived from IFSAR is prone to missing values due to radar geometry effects which cause certain points not to be mapped. Missing data is replaced with the median value of a local neighborhood of user-definable size. We were working with data on a one-foot resolution grid, and a  $5 \times 5$  window produced reasonable results. After one application of median filtering, those pixels which still have no data are replaced with the lowest available height value, computed over the entire image. For convenience, the data is shifted so that the minimum height value is zero.

In order to obtain a smooth segmentation, the noisy elevation data is smoothed prior to thresholding. The amount of smoothing should be sufficient to suppress noise in homogeneous areas, but not to blur edges. Of the various edge-preserving smoothing schemes, the Sigma filter [22] is simple to implement and has been used for removing speckle noise in SAR imagery. According to this filter, the de-noised estimate of a pixel value is the average of those pixels in its neighborhood which are within a two-sigma interval of its value. If the number of such pixels is smaller than a minimum value, the center pixel is replaced with the average of its 8-connected neighborhood. We used a  $7 \times 7$  neighborhood window with a threshold of four for the minimum required number of pixels in the two-sigma range.

**Local thresholding:** The image is divided into non-overlapping square blocks, each of which are thresholded individually. The block size is chosen to be smaller than the typical separation between two buildings, yet large enough to produce a meaningful height histogram. For our high resolution data, a block size of  $16 \times 16$  worked well.

Choice of the specific thresholding scheme within each block was driven by the nature of the data. Although only pixels belonging to two classes were expected to be present in each block, the distribution of pixels within each class was very random. Kittler and Illingworth [16] employed the strategy of assuming that the empirical data-distribution is a mixture of two Gaussians and selecting the threshold to minimize the Kullback distance from the observed histogram to the unknown mixture. This assumption suited our data and therefore we used the corresponding threshold selection scheme.

In order to produce a binned histogram for each block, the floating-point elevation data was rounded off to the nearest integer. While forming the histogram in a block, the pixels in a guard band (of size equal to half the block width) are also included [1]. This ensures that small regions of one class, produced due to the artificial block-boundary, are not overwhelmed by the larger region, while selecting a histogram-based threshold. But the computed threshold is applied only to the pixels in the original block.



After computing the threshold, a test is made to check for the homogeneity of the current block. A block is marked homogeneous if any of the following conditions hold:

- The entropy of the entire block (computed by fitting a single Gaussian model) is lower than the combined entropy of the two clusters that result after thresholding. The mean elevation of the block is stored for future use.
- The means of the clusters produced by thresholding are not well separated. Since the two means represent the average rooftop and ground levels within each block, we used a minimum separation (of three feet) criterion. The mean elevation of the block is stored for future use.
- After thresholding, if the fraction of pixels belonging to any one class is smaller than a threshold (5% in our trials), the block was marked homogeneous. The small region could have resulted due to some noisy pixels within the block. The mean of the larger region was assigned to be the mean of the block.

Pixels in the blocks which fail the test for homogeneity are classified on the basis of the threshold for that block into a ground or rooftop class.

**Re-thresholding of blocks with high thresholds:** It is conceivable that blocks which fall on rooftops, and should have been classified as homogeneous, are thresholded because of the noisy height data on rooftops. The computed thresholds in these blocks tend to be higher than that of their neighbors. As a consequence, along the edge that such a block shares with its neighbor, pixels in the neighboring block are all labeled as rooftop, whereas most of the pixels in the current block are labeled as ground. This produces an artificial gap in the rooftop region for that building.

We handle this problem by re-thresholding the block under test using a histogram of pixels over a larger area. We look at all the non-homogeneous blocks in the 4-connected neighborhood of the current block. If any of those blocks has an unbroken line of pixels labeled as rooftop along the edge it shares with the current block, and less than half the pixels along the same edge in the current block are labeled as rooftop, then the current block is marked for re-thresholding. For each block to be re-thresholded, we find the neighboring block with the lowest threshold among all the neighboring blocks which caused it to be marked for re-thresholding. The assumption here is that since all the neighboring blocks lie on or near the same building, the lowest threshold is good enough for discriminating that building from the ground. The pixels in the two blocks are combined to produce a new histogram, which is then used to re-threshold the current block.

**Classifying homogeneous blocks:** Before attempting to eliminate the artificial boundaries, produced by our local thresholding scheme, homogeneous blocks have to be labeled, in their entirety, as part of rooftop or ground. This classification proceeds in the scan-line order by comparing the mean elevation of a homogeneous block against a threshold value derived from its neighbors' thresholds, and assigning that threshold to the current block:

- If the mean of the current block is above (below) the majority of the thresholds among the

non-homogeneous blocks in its 4-connected neighborhood, it is labeled as rooftop (ground) and is conservatively allocated the maximum (minimum) of the corresponding threshold values.

- If, after the first scan, any block is still marked as homogeneous, the above procedure is repeated with its 8-connected neighborhood.
- Surviving homogeneous blocks are classified according to the maximum threshold among the blocks in its 8-connected neighborhood. This approach is conservative, in the sense that most of these blocks tend to be labeled as ground.

**Re-labeling pixels along building edges:** A connected components algorithm is used to group connected rooftops. Small regions labeled as rooftops (less than 100 pixels) are eliminated and the surviving ones are filled in using morphological processing. The boundaries of buildings tend to be blocky at this stage, due to the artificial edge induced by our block-based thresholding scheme. We refine building edges by reclassifying pixels along the edge, both on and off the building region. A sequence of erosions and dilations performed independently on each building region is used to define regions of confirmed rooftops and ground, as well as ambiguous regions along building boundaries. Since more errors are expected when entire homogeneous rooftop blocks lie along the building boundary, these areas are eroded for up to half the block width. All other intra-block edges are deemed reliable and only a dilation (for up to half the block width) is performed in those blocks. Pixels which are still labeled as rooftop(ground) after the morphological operations are retained as confirmed rooftop (ground) regions. The remaining pixels are reclassified by: looking for a minimum number of sample pixels from rooftop and ground in an increasing local neighborhood; computing the average rooftop/ground elevation in the local neighborhood; and assigning them to one of the two classes based on a Euclidean distance criterion. Finally, another set of morphological operations is performed to eliminate small regions and fill in the surviving building regions.

**Removal of anomalous buildings:** Although unnaturally high thresholds in blocks on building rooftops have been handled by re-thresholding those blocks, unnaturally low thresholds resulting in false buildings has not been handled so far. These arise due to clusters of missing or noisy data in blocks which are in predominantly ground regions. The resulting block-threshold is lower than the one associated with a true building, but nevertheless results in a splitting of pixels within that block into two classes. The resulting buildings have a lower mean elevation than the surrounding ground. Ideally, they can be eliminated by comparing the mean building elevation with the mean elevation of the ground immediately surrounding it. In order to reduce computation, we use an approximation for the mean-ground-elevation estimation. The image is divided into non-overlapping blocks, larger than the ones used for local thresholding ( $64 \times 64$ , in our experiments). The mean ground height is computed in each block. The mean ground elevation immediately around a building is estimated as the weighted sum of the mean ground elevations in all the macro-blocks occupied by that building, the weights being the fraction of pixels of the building which lie in that macro-block.

## 5.2 Experimental results

Experiments in building detection were performed on elevation data on a  $1' \times 1'$  resolution grid, derived from a Sandia Labs' IFSAR. Figure 6(a) shows a sample elevation map, converted to an intensity image, clearly illustrating its noisy nature. Figure 6(b) shows the result of building detection using a manually optimized global threshold of forty-one. The buildings tend to merge with background or other buildings, and besides, there is no automatic way of extracting this threshold value. Figure 6(c) shows the result of building extraction using our local histogram-based thresholding scheme. Finally, Figure 6(d) shows a 3-D rendering of the detected buildings, assuming flat roofs and ground. The anomalous building (bottom right of Figure 6(d)) is not eliminated by our algorithm possibly because of image-edge effects. Figure 7 shows another example of IFSAR-derived DEM imagery. The results of building detection using our local thresholding scheme and the rendered 3-D building image are also shown.

## 6 Conclusions

Automatic interpretation of SAR data with the intent of detecting man-made structures, such as buildings, is a difficult problem. The radar signature of buildings depends on a variety of factors, including, but not limited to, imaging geometry, external shape and internal composition, the presence of nearby objects, and system noise. The possible variability in these factors makes the task of even a seasoned image analyst difficult. But, due to increasing popularity of SAR as an aerial sensor and the sheer quantity of data becoming available, there is a need to develop automatic algorithms for SAR image analysis.

In the absence of any other information from auxiliary sensors, we have tried to develop a building detector which looks for typical characteristics of buildings in high-resolution radar images, namely cardinal streaks and shadows. After repeated trials with a variety of rural and urban images, we have concluded that our method works reasonably well in not-too-cluttered imagery. It may be argued that other objects in radar imagery produce the kind of characteristic features that we have used to detect buildings. Elongated vehicles produce a bright streak followed by a shadow region. But incorporating a minimum projected rooftop width for a valid building, circumvents this problem. We do have erroneous detections in situations where a dark region (due to tree-shadow, water-body, or road) is present at some distance from railroad tracks. Sometimes buildings are missed when their cardinal streak is suppressed or occluded by other buildings or trees. But under such circumstances, even a trained human analyst has difficulty in identifying the building.

A better delineation of buildings is possible if auxiliary information, such as an overlapping EO image or a DEM of the site, is available. One feasible scenario is that the same sensor makes several passes over the site, collecting images from different viewpoints. Although it is possible to use multi-pass images for consolidating the evidence of buildings, we have addressed the problem of building-height extraction from multiple SAR images. As a necessary preliminary step, we have



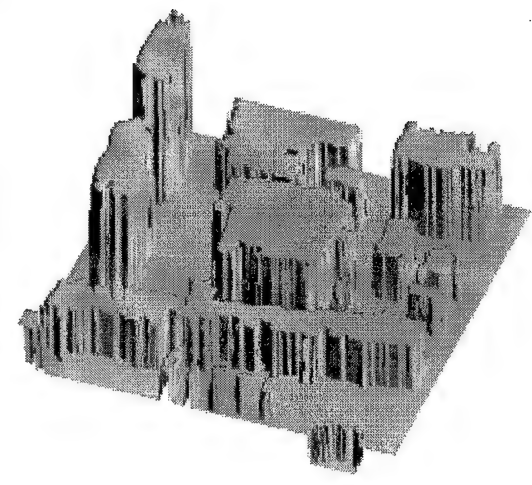
(a)



(b)



(c)

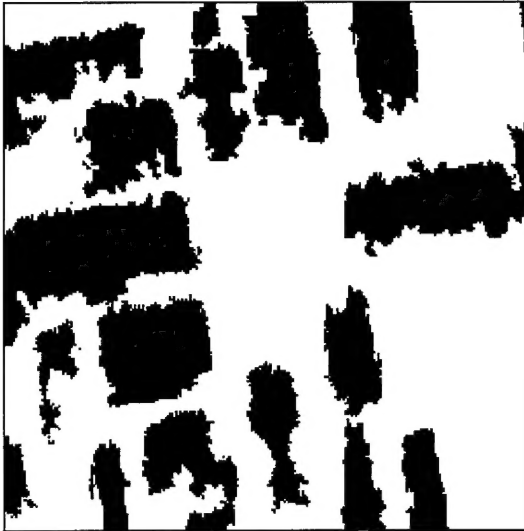


(d)

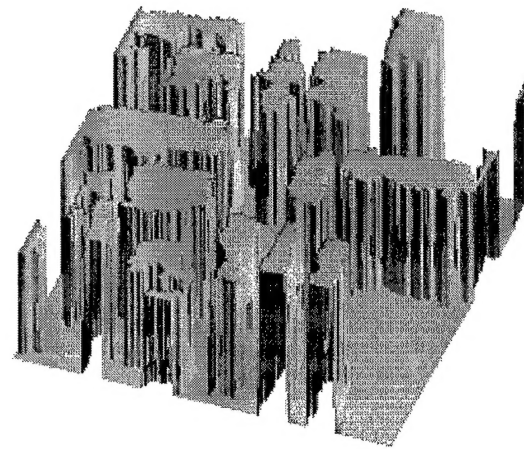
Figure 6: Building detection in IFSAR elevation data: (a) Elevation data rendered as an image, (b) Detected buildings using a global threshold of 41.0, (c) Detected buildings using the local thresholding scheme, (d) 3-D rendering of detected buildings with flat roofs



(a)



(b)



(c)

Figure 7: Another example of building detection in IFSAR-derived elevation data: (a) Rendered image, (b) Detected buildings using the local thresholding scheme, and (c) 3-D rendering of detected buildings with flat roofs



presented a framework for registering multiple airborne SAR images and extracting the heights of linear 3-D structures in them. The accuracy of the estimated heights can be increased if more than two images are available and the building is imaged in all of them.

Finally we have looked at an algorithm for detecting buildings in IFSAR-derived elevation data. High resolution DEMs from IFSAR are slowly becoming available and this report describes one possible approach to exploiting them. Future work along this direction will focus on fitting walls to the building edges, and hypothesizing the shapes of roofs. Although we have merely alluded to it in the context of resolving trees from buildings, the next logical step is to combine IFSAR-derived DEMs and SAR images for building 3-D site models.

## References

- [1] J. R. Beveridge, J. Griffith, R. R. Kohler, A. R. Hanson, and E. M. Riseman. Segmenting images using localized histograms and region merging. *Intl. Journal of Computer Vision*, 2:311–347, 1989.
- [2] D. Canu, J. P. Gambotto, J. A. Sirat, and N. Ayache. Reconstruction of buildings from multiple high resolution images. In *Proc. IEEE Intl. Conf. on Image Processing*, pp. 621–624, Lausanne, Switzerland, September 1996.
- [3] J. Curlander and R. McDonough. *Synthetic Aperture Radar: Systems and signal processing*. Wiley series in remote sensing. John Wiley, New York, 1991.
- [4] H. A. David. *Order Statistics*. John Wiley, New York, 1981.
- [5] H. Derin, P. A. Kelly, G. Vezina, and S. G. Labitt. Modeling and segmentation of speckled images using complex data. *IEEE Trans. Geoscience and Rem. Sensing*, 28(1):76–87, 1990.
- [6] R. E. Fayek and A. K. C. Wong. Extracting buildings from aerial topographic maps. In *Proc. IEEE Intl. Conf. on Image Processing*, pp. 401–404, Lausanne, Switzerland, Sept. 1996.
- [7] H. M. Finn and R. S. Johnson. Adaptive detection mode with threshold control as a function of spatially sampled clutter-level estimates. In *RCA Review*, 29(3):414–464, 1968.
- [8] M. Fischler and R. Bolles. Image understanding research at SRI International. In *Proc. Image Understanding Workshop*, pp. 15–34, Palm Springs, CA, February 1996.
- [9] V. S. Frost and L. S. Yurovsky. Maximum likelihood classification of Synthetic Aperture Radar imagery. In *Computer Vision, Graphics, and Image Processing*, 32:291–313, 1985.
- [10] V. S. Frost, J. A. Stiles, K. S. Shanmugan, and J. C. Holtzman. A model for radar images and its application to adaptive filtering of multiplicative noise. In *IEEE Trans. Pattern Analysis and Machine Intelligence*, 4(2):157–166, 1982.
- [11] J. W. Goodman. Some fundamental properties of speckle. *Journal of the Optical Society of America*, 66(11):1145–1150, 1976.

- [12] M. Herman and T. Kanade. Incremental reconstruction of 3D scenes from multiple, complex images. *Artificial Intelligence*, 30(3):289–341, 1986.
- [13] A. Huertas and R. Nevatia. Detecting buildings in aerial images. *Computer Vision, Graphics and Image Processing*, 41(2):131–152, 1988.
- [14] R. Irvin and D. McKeown. Methods for exploiting the relationship between buildings and their shadows in aerial imagery. *IEEE Trans. on Systems, Man and Cybernetics*, 19(6):1564–1575, 1989.
- [15] C. Jaynes et al. Three-dimensional grouping and information fusion for site modeling from aerial images. In *Proc. Image Understanding Workshop*, pp. 479–490, Palm Springs, CA, February 1996.
- [16] J. Kittler and J. Illingworth. On threshold selection using clustering criteria. *IEEE Trans. on Systems, Man and Cybernetics*, 15(5):652–655, 1985.
- [17] J. A. Kong, A. A. Swartz, H. A. Yueh, L. M. Novak, and R. T. Shin. Identification of terrain cover using the optimum polarimetric classifier. *Journal of Electromagnetic Waves and Applications*, 2(2):171–194, 1987.
- [18] S. Kuttikkad and R. Chellappa. Building wide area 2-D site models from high resolution fully polarimetric Synthetic Aperture Radar images. In *Proc. Intl. Symposium on Computer Vision*, pp. 389–394, Coral Gables, FL, November 1995.
- [19] S. Kuttikkad, R. Chellappa, and L. M. Novak. Building wide area 2-D site models from single- and multipass single-polarization SAR data. In *Proc. SPIE: Algorithms for Synthetic Aperture Radar Imagery III*, vol. 2757, pp. 34–44, Orlando, FL, 1996.
- [20] S. Kuttikkad, R. Meth, and R. Chellappa. Registration and exploitation of multipass airborne Synthetic Aperture Radar images. In *Tech. Report, Center for Automation Research, CAR-TR-857*, April 1997.
- [21] F. W. Leberl. *Radargrammetric Image Processing*. Artech House, Inc., Norwood, MA, 1990.
- [22] J. S. Lee. Speckle suppression and analysis for Synthetic Aperture Radar images. *Optical Engineering*, 25(5):636–643, 1986.
- [23] C. Lin, A. Huertas, and R. Nevatia. Detection of buildings using perceptual grouping and shadows. In *Proc. Computer Vision and Pattern Recog.*, pp. 62–69, Seattle, WA, June 1994.
- [24] C. Lin and R. Nevatia. Building detection and description from monocular aerial images. In *Proc. Image Understanding Workshop*, pp. 461–468, Palm Springs, CA, February 1996.
- [25] Y. Liow and T. Pavlidis. Use of shadows for extracting buildings in aerial images. *Computer Vision, Graphics and Image Processing*, 49(2):242–277, 1990.
- [26] J. McGlone and J. Shufelt. Projective and object space geometry for monocular building extraction. In *Proc. Computer Vision and Pattern Recog.*, pp. 54–61, Seattle, WA, June 1994.

- [27] R. Mohan and R. Nevatia. Using perceptual organization to extract 3-D structures. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 11(11):1121–1139, 1989.
- [28] S. Norohna and R. Nevatia. Detection and description of buildings from multiple aerial images. In *Proc. Image Understanding Workshop*, pp. 469–478, Palm Springs, CA, February 1996.
- [29] E. Rignot and R. Chellappa. Segmentation of polarimetric Synthetic Aperture Radar data. *IEEE Trans. Image Processing*, 1(3):281–300, 1992.
- [30] H. Rohling. Radar CFAR thresholding in clutter and multiple target situations. *IEEE Trans. on Aerospace and Electronic Systems*, 19(4):608–621, 1983.
- [31] M. Roux and D. McKeown. Feature matching for building extraction from multiple views. In *Proc. Computer Vision and Pattern Recognition*, pp. 46–53, Seattle, WA, June 1994.
- [32] V. Venkateswar and R. Chellappa. A framework for interpretation of aerial images. In *Proc. Intl. Conference on Pattern Recognition*, pp. 204–206, Atlantic City, NJ, June 1990.
- [33] U. Weidner and W. Forstner. Towards automatic building extraction from high-resolution digital elevation models. *ISPRS J. of Photogrammetry and Rem. Sensing*, 50(4):38–49, 1995.
- [34] S. H. Yueh, J. A. Kong, J. K. Jao, R. T. Shin, and L. M. Novak. K-distribution and polarimetric terrain radar clutter. *J. of Electromagnetic Waves and Applications*, 3:747–768, August 1989.
- [35] H. Zebker and R. Goldstein. Topographic mapping from interferometric SAR observations. *Journal of Geophysical Research*, 91(B5):4993–4999, 1986.